Towards Measuring Maturity of Privacy-Enhancing Technologies

Marit Hansen[1], Jaap-Henk Hoepman[2], and Meiko Jensen[1]

[1] Unabhängiges Landeszentrum für Datenschutz, Kiel, Germany
marit.hansen@privacyresearch.eu and meiko.jensen@rub.de
[2] Radboud University, Nijmegen, The Netherlands
jhh@cs.ru.nl

Abstract. The assessment of the maturity of Privacy-Enhancing Technologies (PETs) is a complex and challenging task, which can only be performed by experts in the field. However, at the same time, the need for precise technology readiness and quality definitions for PETs emerges rapidly. In order to overcome this gap, standardised means to assess, discuss, and compare PET maturity levels are necessary.
In this paper, we propose an approach for assessing the maturity of PETs. We define both the scales and the methodology for measuring maturity of PETs, in a way that is independent from the domain of application. Based on an in-depth analysis of the criteria to be met by such a PET maturity level scheme, we propose a combined quality-and-readiness level scale to be used for this purpose.

1 Introduction

Since decades, the idea of incorporating privacy and data protection criteria in the design of systems has been discussed. Early work on confidentiality (e.g. based on cryptographic algorithms) or anonymity and pseudonymity (e.g. Mix networks) showed that technology can support or even ensure privacy and data protection features. A special category of technologies that aimed at enhancing privacy was coined "Privacy-Enhancing Technologies (PETs)" [14].

Recent and upcoming legal norms demand "privacy by design" (the European Data Protection Regulation [2] as well as the recently passed eIDAS Regulation [3]). However, how to transpose this into the system design process is either not detailed or left to secondary legislation such as delegated or implementing acts. The ENISA report on Privacy and Data Protection by Design [8] gives an overview on today's landscape concerning privacy engineering. PETs are recognised as an important element in the overall design task. The ENISA report points out that the solutions, techniques, and building blocks presented are of differing maturity levels—without providing criteria on how to assess the individual maturity.

In this paper, we specify these criteria for the first time, and take the first steps towards defining a full-fledged PET maturity assessment methodology, based on existing work in other fields of technology (e.g. NASA's scale of technology readiness levels (TRLs), [20]).

One crucial finding in our work is the strong belief that a mere assessment of technology readiness may yield misleading results. More precisely, a PET that is available and deployed, but shows severe shortcomings concerning its quality regarding privacy protection, should not be preferred over a better privacy technology that—perhaps because of the predominance of the worse technology—scores lower on the readiness scale. For this reason, we decided to pursue a two-fold strategy that tackles technology readiness as one dimension and privacy enhancement quality as a second dimension. The individual results then are combined into a single PET maturity score.

We aim to ensure that the assessment scheme for PET maturity we are developing is useful for a diverse set of potential stakeholders, such as Data Protection Authorities (DPAs), data controllers and data processors, developers, certification bodies, auditors, or standardisation bodies. The relevance of PET maturity for this diverse set of stakeholders demands that the information has to be easily comprehensible by experts and laypeople; potential misinterpretation of the information should be prevented as far as possible. Moreover, the methodology has to be adaptable to all kinds of PETs (e.g. protocols, algorithms, software, hardware, products, IT-based services; ideas, concepts, specifications, implementations, workable demonstrations, rolled-out versions, etc.).

The text is organised as follows: Section 2 introduces important terms and notions that are necessary to determine the scope of the project. In particular, the term *Privacy-Enhancing Technology (PET)* will be discussed. A survey of existing methods to measure technology readiness is given in Section 3. Our proposal for a PET maturity scale based on both a readiness level and a quality level is presented in Section 4. A first sketch of the corresponding methodology to score a given PET on the defined maturity scale is presented in Section 5. Finally, Section 6 summarises our findings and gives an outlook on our intended future work.

2 Setting the Stage

In this section, we introduce the basic terminology used throughout the paper, and the underlying concepts and related work we base our proposal on. We also point out some gaps and pitfalls with respect to the semantics of certain commonly used terms, and clarify how we interpret them.

For example, in this paper we distinguish levels and scales as follows. A *level* is the particular score on a metric, e.g. `pilot` as the value for the readiness level. A *scale* is the set of levels a certain metric can assume. An *indicator* is a factor that may be meaningful for determining the level; it is an input value for the assessment. *Evidence* denotes the set of indicators.

2.1 Privacy-Enhancing Technology

Privacy-enhancing technologies (PETs) have been characterised in various ways. Some authors [6] define them quite specifically as "a coherent system of ICT measures that protects privacy by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data". The OECD Report on PETs from 2002 [23] takes a broader perspective and also declares tools "that allow a user to choose if, when and under what circumstances personal information is disclosed" in scope. The European Commission [10] considers a wider range of PETs that include those that support legal compliance with data protection regulation.

For assessing PET, we aim at allowing a wide definition of PETs, encompassing all kinds of technologies that support privacy or data protection features (e.g. technologies that make use of privacy design strategies [15] or consider protection goals for privacy engineering [13]). Compared to a definition that restricts PETs to data minimisation, this approach provides greater flexibility and adaptability, albeit adds complexity when statements on the privacy enhancement properties in various categories have to be elaborated. Our approach is detailed in Section 5.2.

2.2 The Technology Lifecycle

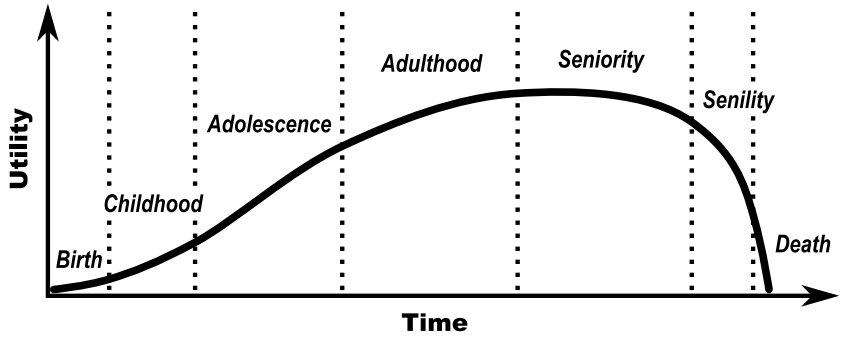

Fig. 1. Lifecycle of a technology (adapted from [21]).

We distinguish between seven different phases within the lifecycle of a technology, illustrated in Figure 1, as defined by William L. Nolte [21]. Initially, each technology starts off with an idea, its *birth*. Then, this idea is analysed preliminarily, elaborated on, and considered useful. Thus, in the next phase, the idea is discussed on a broad scale, e.g. within research and development communities. Yet, there is no working prototype, not even a demonstrator, so the correlated phase is that of *childhood*. At some point, a proof of concept is implemented in test environments under laboratory conditions, marking a progress towards

adolescence level. Depending on the complexity of the technology, the transition from childhood to adolescence can be rapid (e.g. if the idea gets implemented by its inventor straight away) or can take decades (e.g. if the idea cannot be implemented with current state-of-the-art technology).

The next step is that of a real-world usage of the technology under non-laboratory conditions. Typically, this step is performed with the release of a first complete implementation, or with the advent of a pilot implementation in real-world systems. Thus, the technology matures towards a state of *adulthood*.

Subsequently, the next remarkable transition is that of a full market participation of the technology, which is typically kicked off by advent of a ready-to-use product being sold (rented, consulted for, commercially supported for, etc.). This implies that the maturity of the technology has reached a point where it becomes feasible to gain profits from utilising the technology to such extent that a market emerges. The corresponding age is that of *seniority*.

Finally, the technology might become obsoleted by technological evolution. For PETs, this could mean that devastating attack techniques render the technology useless in an irreparable way, or simply by the advent of a superior technology that provides the same guarantees in a more favorable way. In each of these cases, the use of the technology decreases (into what we may call the *senility* phase), until it fades out of use, and reaches its final state of *death*.

This lifecycle model has been used as the basis for our readiness metric defined in Section 4.1.

2.3 What Makes a Scale Effective?

The effectiveness of a scale depends its *comprehensibility*, its *comparability*, its *scorability*, and its *reproducibility*. We define these four criteria as follows:

Comprehensibility First of all a score should be easy to understand and to apply by users looking for an appropriate PET to solve a particular problem in a certain context[3]. The meaning of a certain score should be intuitively clear.

Comparability Similarly, comparing different results should be straightforward. It is especially important to know for combined scores resulting from different dimensions (readiness and quality) whether — and under which conditions — comparability is given.

Scorability Further, a particular PET should be easy to score objectively on the scale at hand by an evaluator. The score should be derived from clearly described indicators, that are easy to determine or measure for an arbitrary PET that is going to be evaluated. Moreover it should be clear how a combination of values or appreciations for the different indicators should be combined into the overall score.

Reproducibility Finally, a score for a PET on some scale should be reproducible. This means that a PET should receive (almost) the same score, when independently scored by two or more evaluators. This further emphasises the objectiveness implicit in the definition of scorability.

[3] We note that in our methodology the application context of a PET is out of scope for determining its maturity, as explained further on in this report.

3 Related Work

Since we regard maturity of PETs as a combination of their readiness and their privacy enhancement quality, we have to consider related work from both fields.

Technology Readiness Levels (TRLs) have been used for about 40 years [22] especially by NASA ([20]) and in the military sector. They are based on a nine-point score (TRL 1-9) where lower TRLs express early development and readiness stages while high levels denote completely developed and thoroughly tested systems.

Similarly, the European Commission has introduced a similar nine-point scale for technology readiness for the work programme 2014-2015 (Horizon 2020) [11]—with similar advantages and disadvantages:

- TRL 1: basic principles observed
- TRL 2: technology concept formulated
- TRL 3: experimental proof of concept
- TRL 4: technology validated in lab
- TRL 5: technology validated in relevant environment (industrially relevant environment in the case of key enabling technologies)
- TRL 6: technology demonstrated in relevant environment (industrially relevant environment in the case of key enabling technologies)
- TRL 7: system prototype demonstration in operational environment
- TRL 8: system complete and qualified
- TRL 9: actual system proven in operational environment (competitive manufacturing in the case of key enabling technologies; or in space)

For being able to assess the readiness of a system, the evaluation process can be supported by a TRL Assessment Matrix and tools such as a TRL Calculator as developed for the NASA TRL scheme [5]. In the beginning, TRLs were mainly assigned to developed hardware; later, software or combined systems were taken into account, too.

Since its publication, the TRL scale has been discussed and criticised, in particular by pointing out limitations and needs for a multidimensional approach (e.g. [21]). Also for derived scales such as a Systems Readiness Level (SRL) (cf. [24]) it is being heatedly debated whether they are misleading and may be dangerous because of arbitrary assessment results, and how potential problems could be overcome (cf. [17]). Here it became evident that readiness should be understood in context and that it is usually not sufficient to assess "readiness" without regarding "quality" [25].

In the context of privacy and security this additional quality dimension is especially viable because there are many examples of widely deployed technology (that would score high on a pure "readiness" scale) that provide sub-optimal protection in practice.

In this respect, standards for software quality such as ISO/IEC 25010 on Systems and Software Quality Requirements and Evaluation (SQuaRE) [16] or, where applicable, for process quality such as ISO/IEC 15504 on Software Process Improvement and Capability Determination (SPICE) [1] have to be considered.

However, these standards are not comprehensive, but extensions are possible (e.g. shown in [12] for extending SQuaRE by green and reliability issues). Other criteria may be more or less neglected for assessment of PET maturity since they most likely won't play a role.

Moreover, measurement of privacy enhancement quality is not a trivial task. Since this is not the focus of this paper, we only mention some noteworthy contributions that may provide some input to a PET maturity debate, among others the work on comparing different degree of anonymity (e.g. concerning differential privacy [9], k-anonymity [27], l-diversity [19], or t-closeness [18]) or on calculations of linkability (e.g. [7, 4]).

4 PET Maturity Metric

We are now ready to define our PET maturity metric. We will do so by defining our scale for readiness, followed by our definition of a quality scale, and continuing by describing how scores on both scales are combined to obtain the overall PET maturity level. Further, we analyse the tensions between measurable indicators and expert opinions.

4.1 A Scale for Readiness

We begin by defining a scale along which to express the *readiness* of a certain PET inline with the phases of the technology lifecycle described in Section 2.2. Readiness of a PET expresses whether a PET can be deployed in practice at a large scale, or that it can only be used within a research project to build upon to advance the state of the art in privacy protection. Readiness also says something about the amount of effort (in terms of time and money) still needed to allow the PET to be really used in practice. To ensure comprehensibility (see Section 2.3), we choose not to score readiness by a simple number on some linear scale. Instead we define the following readiness levels for a PET.

`idea` Lowest level of readiness. The PET has been proposed as an idea in an informal fashion, e.g. written as a blog post, discussed at a conference, described in a white paper or technical report.

`research` The PET is a serious object of rigorous scientific study. At least one (but preferably more) serious academic paper(s) have been published in the scientific literature, discussing the PET in detail and at least arguing its correctness and security and privacy properties.

`prototype` The PET has successfully been implemented, and can be tested for performance and other properties in practice. "Running code" is available.

`pilot` The PET is or has (recently) been used in some small or larger scale pilot applications with real users. The scope of application, and the user base may have been restricted (e.g. to power users, students, etc.).

`product` The highest readiness level. The PET has been incorporated in one or more generally available products that have been or are being used in practice by a significant number of users. The user group is not a priori restricted (by the developers).

`outdated` The PET is not used anymore, e.g., because the need for the PET has faded, because it is depending on another technology that is not maintained anymore, or because there are better PETs that have superseded that PET.

These readiness levels relate to the technology lifecycle; a later evolutionary level does not necessarily mean that the PET is better, because the aging process may not improve the PET's maturity or its applicability when it becomes `outdated`. This readiness level indicates that the PET should no longer be used. The transition from one readiness level to the next is not as sharply delineated as the previous scale suggests. In fact, different PETs that belong to the same readiness level may differ significantly. Some barely made it the level assigned to them; others are about to enter the next level. To allow people to express these differences, a readiness level may be augmented with the next higher readiness level in the scale above. So, for example, a readiness level of `pilot/product` may be appropriate for a PET that has been used in several pilot programmes and is currently being beta-tested as a (commercial) general purpose product.

4.2 A Scale for Quality

Although *quality* is somewhat dependent on readiness (a rolled out product has received so much more attention over the years than a concept still in its research stage), the quality of a PET is not only determined by its readiness. In fact several PETs at the same readiness level may have varying levels of quality. As argued in the introduction, it is important to realise that sometimes a PET with high readiness may still have a low quality. We now turn to make this notion of quality more precise.

We base our approach on the ISO/IEC system and software quality models ISO standard 25010 [16], but adjust and refine it to our needs. ISO 25010 distinguishes the following eight quality characteristics: 'functional suitability', 'reliability', 'operability', 'performance efficiency', 'security', 'compatibility', 'maintainability' and 'transferability'. Not all of these characteristics are relevant for our purposes. Some characteristics are more important than others and therefore contribute more to the overall quality score.

For example, because we want the overall maturity scale to be independent of the particular context in which a PET is applied, characteristics like functional suitability are out of scope. We believe that a PET with limited functionality has the same quality as one with a larger (or different) functionality. Which one to choose depends on the requirements to be met within a particular application context.

Similarly, 'compatibility' is deemed a less relevant characteristic.

We interpret 'operability'—which refers to the degree to which a product is easy to learn, understand, and attractive to a user—to be directed at a system developer instead of an ordinary user (because a PET is typically embedded into larger system, and not directly exposed to the user).

The 'security' characteristic is renamed to 'protection', and focuses on preventing privacy infringements. A separate characteristic 'trust assumptions' is

added to capture whether and if so how much trust in certain components and agents is assumed.

Also added are two other characteristics: 'side effects' and 'scope'. This brings us to define the quality scale as comprising the following nine PET quality characteristics, listed in decreasing order of importance

protection The degree of protection offered (in terms of for example unlinkability, transparency, and intervenability) to prevent privacy infringements while allowing access and normal functionality for authorised agents. Also depends on the type of threats and attacks against which the PET offers protection.

trust assumptions The number of components and/or agents that need to be trusted, and the nature and extent of trust that must assumed in order to use the PET. Also depends on whether these assumptions are legal, organisational, procedural, or technical.

side effects The extent in which the PET introduces (undesirable) side effects. Measured in terms of composability.

reliability The degree to which a system or component performs specified functions under specified conditions for a specified period of time. Measured in terms of fault tolerance, recoverability, and compliance. Also measured in terms of the number of vulnerabilities discovered.

performance efficiency The performance relative to the amount of resources used under stated conditions. Measured in terms of resource use (storage, CPU power, and bandwidth) and speed (latency and throughput).

operability The degree to which the product has attributes that enable it to be understood, and easily (and in particular securely) integrated into a larger system by a qualified system developer. Measured in terms of appropriateness, recognisability, learnability, technical accessibility, and compliance.

maintainability The degree of effectiveness and efficiency with which the product can be modified. Measured in terms of modularity, reusability, analysability, changeability, modification stability, and testability. Open source software typically scores high on this characteristic. Also, systems that have an active developer community, or that have official support, score high.

transferability The degree to which a system or component can be effectively and efficiently transferred from one hardware, software or other operational or usage environment to another. Measured in terms of portability and adaptability.

scope The number of different application domains the PET is applied in or is applicable to.

Usually, each of these characteristics is relevant for a PET, independent of its readiness level. However, the indicators that determine the score for each of the characteristics *do* depend on the readiness level. For example, the quality of a rolled out product depends on how well it is supported (by a help desk, code updates, etc.). These indicators are irrelevant for research level PETs. The quality of those is determined more by the ranking of the venues in which the research is published, for example.

For each of these nine characteristics, a PET can receive a score in the range

−− (very poor)	− (poor)	0 (satisfactory)	+ (good)	++ (very good)

The overall quality level also utilises this five-value scale, and is comprised of the nine individual scores, according to a specific quality evaluation function, as discussed in Section 5.5.

4.3 Combining Readiness and Quality to Express Maturity

The scales for readiness and quality defined above allow us to define the real scale we are interested in: a scale for PET maturity. In fact this overall scale is simply the combination of the readiness level superscripted by the quality level.

$$\texttt{readiness}^{\text{quality}}$$

So for example a PET with readiness level `pilot` and quality + has an overall PET maturity level of $\texttt{pilot}^{+}$. Thus, the total set of potential PET maturity values spans from $\texttt{idea}^{--}$ and $\texttt{idea}^{++}$ to $\texttt{outdated}^{--}$ and $\texttt{outdated}^{++}$.

4.4 Evidence: Measurable Indicators vs. Expert Opinions

When assessing maturity of a PET, different experts may have different opinions with respect to its readiness and quality. Hence, each assessment approach that is solely based on expert opinions is likely to be affected by the choice of experts, and thus lacks reproducibility. Having the same PET assessed by different expert groups may lead to different assessment results, due to the different viewpoints and discussion dynamics among the chosen sets of experts.

In order to mitigate this biased assessment approach, it needs to have some indisputable parameters to be taken into account. Such parameters should be assessable in a way that is unambiguous, leading to the exact same parameter value and assessment indication no matter who performs the parameter assessment. We call these types of parameters *measurable indicators*, meaning that they indicate an assessment result based on objective evidence. As such, measurable indicators are robust against change of assessors, as different assessment instances of the same measurable indicator will always result in the same indicator values, and thus in the same assessment result.

Examples for potential measurable indicators in the field of PET maturity assessment are:

- number of scientific publications referring to the PET;
- number and type of audits/certifications performed for the PET;
- number of university courses covering the PET topic;
- number of similar products in the market if the PET is a product;
- number of hits when searching for the PET in online search engines; or
- number of years since the PET was initially proposed.

As can be seen, each of these measurable indicators represents a certain characteristic with respect to the PET, and does so in an indisputable way. There can be no two different opinions on the total number of scientific publications referring to the PET, for example, at least not on a level of significance. Such a value is an objective evidence for a certain level of maturity of a PET.

However, though assessing these measurable indicators is feasible and quite robust, determining its implications with respect to the result of the assessment is more challenging. What does the number of search engine hits for the PET say about the maturity of a PET? What should be the impact of the existence of six different privacy certifications of a PET product? Each of these measurable indicators gives a small implication on the level of maturity the PET has probably reached. For instance, the existence of a substantial amount of competing products in the market of the PET to be assessed clearly implies that this PET has reached at least the `pilot` stage, more likely even the `product` stage of readiness. If there are no products in the market at all, this might indicate an earlier maturity stage, probably `research`, but it might also be the case that the PET itself is not suitable to be sold as a dedicated product. Nevertheless, it still could be utilised in many products out there, and still could be in the `product` readiness stage.

The measurable indicators are robust in assessment, but fuzzy in their implications to the result of the assessment. They need to be included in the overall assessment process, in order to mitigate the impact of assessor choices, but they are not precise enough to be used as the only, not even as the major base for a PET maturity assessment. Thus, we propose to utilise these indicators as input, but combine them with inputs from a dedicated *board of experts*.

5 The Assessment Process

Based on the findings described in the previous sections, we outline the process of performing a PET maturity assessment. This five-step approach is explained in the following subsections.

5.1 Overview

The process of assessing PET maturity along the lines defined in this document involves five steps, as illustrated in Figure 2. The implicit initial step of an assessment consists in the determination of the assessor, as that is a very critical entity in performing the assessment. The role of the assessor is that of an expert in performing assessments. Beyond that, expertise both in terms of privacy and in the domain of interest the PET is assessed in would be beneficial. Moreover, the assessor needs to be unbiased, as far as possible, and objective in all decisions.

In the first explicit step of the assessment, it is necessary to select and precisely define the *Target of Assessment*, i.e. the concept, technology, or product that is to be assessed. Details on this step are given in Section 5.2.

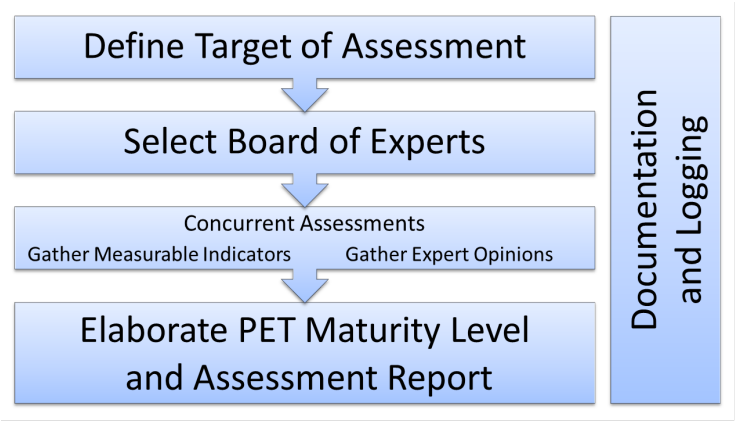

Fig. 2. Overview of the PET Maturity Assessment Process

Once the Target of Assessment is defined, the next step consists in gathering the board of experts to be asked for their opinion. Ideally, the experts should have expertise both in the domain of application the PET is evaluated for, and in the privacy engineering discipline. As with the assessor, it is necessary to gather an unbiased, objective, heterogeneous set of experts for this task (cf. Section 4.4), as far as this is feasible. Though there is no upper bound on the number of experts, we propose a minimum of five experts to be involved in the board.

This step also concludes the preparation phase of the assessment.

Comprising the major part of the assessment, the next two steps can be performed in parallel. On the one hand, it is necessary to gather a specific set of scores to be evaluated from public information sources. For instance, this may cover tasks such as counting the number of research publications that refer to a given PET, or similar assessment of objective indicators with respect to maturity (that is both readiness and quality) of the PET in consideration. This step would typically be performed by the assessor.

Concurrently, and somewhat independent from the previous step, the board of experts needs to be asked for their opinion with respect to the PET in consideration.

Once both concurrent steps are completed, and the total set of evidence gathered for this assessment is compiled, the final and most critical step consists in the aggregation of the assessment results. Performed by the assessor, this step involves three tasks:

1. determination of the level of technology readiness of the PET, according to the scale defined in Section 4.1,
2. assessment of the overall quality of the PET, according to the quality characteristics described in Section 4.2, and
3. aggregation of these two intermediate assessments into the final PET maturity level, as discussed in Section 4.3.

Finally, the documentation and logging inputs, which were collected throughout the other steps of the assessment, need to be aggregated, and comprise a PET Maturity Assessment Report accompanying the PET maturity level achieved.

Once the final PET maturity result is obtained, and the PET Maturity Assessment Report is completed, the assessment process concludes.

5.2 Defining the Target of Assessment

The initial step of assessing a given PET's level of maturity is the precise definition of the *Target of Assessment (ToA)*. Depending on its phase in the technology lifecycle as outlined in Section 2.2, a PET may consist of a few lines of demonstrator source code only, or may already have been implemented in a set of software products being sold and bought in a dedicated market of its own. Thus, the defining the correct ToA can be quite tricky.

If a PET is in one of its early stages of evolution, where it merely is made up by a concept outline or a rough set of ideas, the ToA typically consists of the major concept of the PET, as outlined by its maintainers. Being a theoretical concept without even a basic implementation, measurable quantitative maturity indicators like market share, lines of source code, etc., are not available, and thus cannot be used for maturity assessment. Available measurable maturity indicators for this stage of maturity can only be found in the research and discussion domain (such as number of research papers published that refer to this PET).

If a well-maintained implementation of a PET already exists, but no commercially available product along this implementation (such as a software product, consulting services, support desk, or the like) is found in the open markets, the ToA can be narrowed down to the scope of this implementation. Whenever a precise condition of the PET in question is required within the assessment, the concept is evaluated according to the details found in this implementation. Also, measurable maturity indicators from the source code realm (like lines of code, amount of source code documentation, etc.) can be used based on the numbers available for the existing implementation.

If a dedicated market for solutions utilising this PET already is in place, the ToA can no longer be defined as the (single) concept or implementation of the PET. Given that different products and different domains of application may result in differing privacy guarantees, the ToA in this case has to be narrowed down to one of the existing products or implementations only. This is due to the fact that different implementations of the same PET may have different characteristics, different levels of completeness, and different levels of quality. Thus, an assessment should focus on a single product or implementation only, potentially relating it to other products of the same category for comparison, but fixing the ToA on the product, not on the theoretical concept beneath. Measurable indicators for such a level of maturity may range from market share data to sales numbers, active developer community sizes, and total amount of financial capital allocated to utilisation of the PET, among others.

5.3 The Assessment Methodology

As shown in Figure 3, our methodology is based on both the measurable indicators as well as the expert opinions, collected for both readiness and quality assessment. More precisely, the measurable indicators are collected and normalised according to reasonable individual scales, depending on the ToA. The expert opinions are collected by means of dedicated forms, consisting of both a scale-based assessment and a detailed opinion comment part. Then, all of these inputs are processed by the assessor to gather two separate intermediate results: a *Readiness Score* and a *Quality Assessment*. Finally, both of these are combined into the final *PET Maturity Level*.

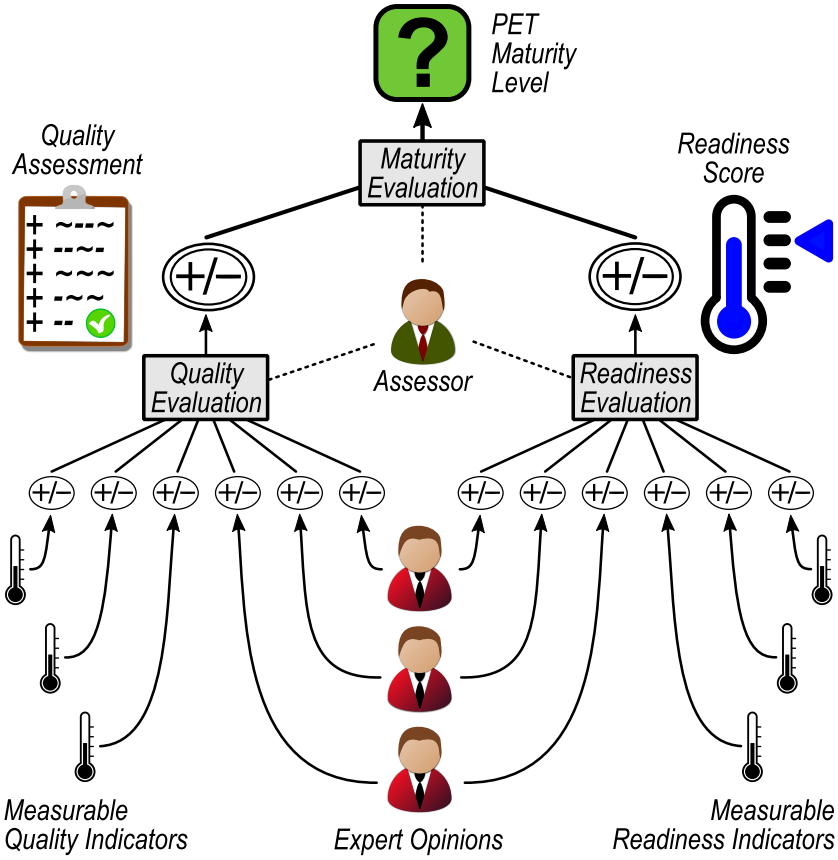

Fig. 3. PET Maturity Assessment Methodology

5.4 Readiness Assessment

The readiness assessment of the ToA begins with the selection and harmonisation of all measurable indicators to be used, a task we propose to be performed by the assessor. The expert opinions for readiness assessment are collected by means of asking each expert on her assumed readiness level of the ToA (ranging from `idea` to `outdated`, as described in Section 4.1), with the option to choose two adjacent levels at once, if the expert thinks the ToA is in a transition from one level to the next.

The next step for the assessor consists in harmonising the results gathered from the initial part of our approach. Regarding the expert's feedback, the assessor needs to identify the dominating level from the votes, but also check for consistency among the total set of responses. A strong deviation of levels may indicate the need for additional discussion and harmonisation among the experts, as it clearly indicated differences in the perception of the ToA among the set of experts. Thus, the assessor needs to verify a certain level of homogeneity of expert opinions before proceeding with the assessment.

Regarding the measurable indicators, the type of ToA already allows for some estimations regarding the set of indicators to consider for readiness level assessment. If the number of ToA-comparable products in the market is large enough, this already gives a clear indicator that the level of `prototype` has already passed. However, the final selection and balancing of measurable indicators to be considered is a task that is always to be performed by the assessor.

As a result, the combination of harmonised expert opinions and measurable indicators makes the final readiness level to be assigned to the ToA.

5.5 Quality Assessment

The main inputs for quality assessment in our approach are the measurable indicators of relevance for quantification of quality, and the expert opinions with respect to the ToA's privacy enhancement quality.

Herein, the measurable indicators may vary depending on the type of ToA. For instance, the number of successful audits or certifications of an existing product as ToA has some indications for its assumed quality, but is obviously not a feasible indicator for a research-stage concept ToA that cannot be audited yet. Thus, the selection and balancing of reasonable measurable indicators for the given ToA is performed by the assessor.

The second input to the Quality Assessment in our approach consists in the dedicated feedback from experts. Each of the experts is asked to answer a few questions with respect to the quality of the ToA in terms of the nine quality characteristics as described in Section 4.2. Each expert is asked to rate the ToA on the quality scale ($--$ to $++$) for each of these nine criteria. Once this process is completed, the assessor evaluates these findings, elaborating the dominating quality characteristics of the ToA. Therein, the assessor may also incorporate findings from the separate comments given by the experts, e.g. in order to spot

domain-specific strengths or weaknesses, or even showstopper arguments against the use of a ToA.

The result of the quality assessment part of our approach is a dedicated Quality Assessment Report, comprising of all expert opinions, including their scores for the nine quality characteristics and comments, and all measurable indicators used in the assessment. This report, which should give a quite decent estimation on the quality of the ToA, can then be used in a later stage to decide upon the final PET Maturity Level, as described in Section 5.6.

5.6 PET Maturity Assessment

The last step in performing a full PET maturity assessment of the ToA consists in combining the results from the quality assessment part with the achieved readiness level. In our approach, this task narrows down to aggregating the Quality Assessment Report's findings into a single quality indicator (on the quality scale described in Section 4.2), and attaching that quality indicator to the readiness level of the ToA. The combined result thus is a bipartite value anywhere in the range between $\texttt{idea}^{--}$ to $\texttt{idea}^{++}$ and $\texttt{outdated}^{--}$ to $\texttt{outdated}^{++}$.

6 Conclusions

6.1 Discussion

The PET maturity metric we propose is independent of the specific context in which the PET is applied. This is different from some technology readiness metrics that explicitly define the readiness of a technology with respect to the particular context in which it is applied (cf. e.g. [25, 26]).

The advantage of our approach is that the maturity of a PET can be scored just by evaluating the PET itself. This makes it easier to assess the maturity of a PET. As a consequence, however, the maturity of a PET by itself does not say whether it is suitable to apply in a certain context. To make that decision, the requirements imposed by the context need to be matched with the functionality, properties, and guarantees as well as potential dependencies or side effects of the PET under consideration.

Our aim is to objectify the assessment of PET maturity, but at the same time we are convinced that a fully automated solution would not produce reliable results. Instead we belief that involvement of (human) experts will be necessary for a meaningful assessment, albeit supported be measurable evidence. Robustness and validity of our approach can only be achieved if an unnoticeable manipulation of the results can be sufficiently prevented. This will highly influence the choice of experts and the measurement methods, but also the transparency of the (final and probably also interim) results of the assessment so that they can be well comprehended by the various target groups, e.g. users, DPAs or funding agencies.

6.2 Future Work and Research Indications

Our PET maturity levels can be utilised in various different scenarios of application. For instance, they can help companies to identify PETs of relevance for their business domain, e.g. for utilisation in existing products. They can be used by funding agencies for identifying interesting PETs that are close to market, in order to provide support for entrepreneurs. DPAs can utilise the PET maturity levels for discussing the legal definition of the technological state of the art.

For all of these domains, the validity and utility of the PET maturity levels need to be thoroughly tested prior to fixation (e.g. by means of standardisation). Thus, our obvious future work consists in choosing and assessing a multitude of PETs with respect to their maturity, and thereby validate both the scale and the methodology of our approach. As this task comes with huge efforts, intense research on means to support, (semi-)automate, and optimise such broad-scale PET maturity assessments becomes necessary.

Acknowledgements. The work of M.H. and M.J. was partially funded by the European Commission, FP7 ICT program, under contract no. 318424 (FutureID project).

References

1. ISO/IEC 15504-5: Information technology – Process assessment – Part 5: An exemplar software life cycle process assessment model. Tech. rep., ISO JTC 1/SC 7 (2012)
2. European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 C7-0025/2012 2012/0011(COD)) (2014) (2014), http://www.europarl.europa.eu/sides/getDoc.do?type=TA&language=EN&reference=P7-TA-2014-0212
3. Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC. OJ L 257, 28.08.2014, pp. 73–114 (2014)
4. Berthold, S.: Inter-temporal Privacy Metrics. Doctoral thesis, Karlstads Universitet (2014), http://kau.diva-portal.org/smash/get/diva2:757291/FULLTEXT01.pdf
5. Bilbro, J.W.: Systematic Assessment of the Program / Project Impacts of Technological Advancement and Insertion Revision A. Tech. rep. (2007), https://acc.dau.mil/adl/en-US/320595/file/46759/White%20Paper%20on%20Technology%20Assessment%20Rev%20A.doc
6. Borking, J.J., Raab, C.D.: Laws, PETs and other Technologies for Privacy Protection. Journal of Information, Law & Technology (JILT) 1(1) (2001), http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2001_1/borking

7. Clauß, S.: A Framework for Quantification of Linkability Within a Privacy-enhancing Identity Management System. In: Proceedings of the 2006 International Conference on Emerging Trends in Information and Communication Security. pp. 191–205. ETRICS'06, Springer-Verlag, Berlin, Heidelberg (2006), http://dx.doi.org/10.1007/11766155_14
8. Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.H., Le Métayer, D., Tirtea, R., Schiffner, S.: Privacy and Data Protection by Design – from policy to engineering. Tech. rep., ENISA (2014), http://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/privacy-and-data-protection-by-design/at_download/fullReport
9. Dwork, C.: Differential Privacy. In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (eds.) Automata, Languages and Programming – ICALP 2006, Lecture Notes in Computer Science, vol. 4052, pp. 1–12. Springer Berlin / Heidelberg (2006)
10. European Commission: Privacy Enhancing Technologies (PETs) – the existing legal framework. MEMO/07/159 (May 2007), http://europa.eu/rapid/press-release_MEMO-07-159_en.htm
11. European Commission: Horizon 2020 – Work Programme 2014-2015, Annex G. Technology readiness levels (TRL). European Commission Decision C (2014)4995 of 22 July 2014. Tech. rep. (2014), http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf
12. Gordieiev, O., Kharchenko, V., Fusani, M.: Evolution of Software Quality Models: Green and Reliability Issues. In: Proceedings of the 11th International Conference on ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer (ICTERI 2015). vol. 1356, pp. 432–445 (2015), http://ceur-ws.org/Vol-1356/paper_71.pdf
13. Hansen, M., Jensen, M., Rost, M.: Protection Goals for Engineering Privacy. In: 2015 International Workshop on Privacy Engineering (IWPE). IEEE eXplore (2015), (to appear)
14. Hes, R., Borking, J.J.: Privacy-Enhancing Technologies: The Path to Anonymity. Tech. rep., Registratiekamer (1995)
15. Hoepman, J.: Privacy Design Strategies (extended abstract). In: ICT Systems Security and Privacy Protection – 29th IFIP TC 11 International Conference, SEC. pp. 446–459 (2014)
16. ISO/IEC 25010: Systems and software engineering – Systems and software quality requirements and evaluation (SQuaRE) – System and software quality models. Tech. rep., ISO JTC 1/SC 7 (2011)
17. Kujawski, E.: Analysis and Critique of the System Readiness Level. IEEE T. Systems, Man, and Cybernetics: Systems 43(4), 979–987 (2013), http://dx.doi.org/10.1109/TSMCA.2012.2209868
18. Li, N., Li, T., Venkatasubramanian, S.: t-Closeness: Privacy beyond k-anonymity and l-diversity. In: Chirkova, R., Dogac, A., Özsu, M.T., Sellis, T.K. (eds.) ICDE. pp. 106–115. IEEE (2007)
19. Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M.: l-Diversity: Privacy beyond k-anonymity. ACM Trans. Knowl. Discov. Data 1(1) (Mar 2007)
20. Mankins, J.C.: Technology Readiness Assessments: A Retrospective. Acta Astronautica 65, 1216–1223 (2009)
21. Nolte, W.L.: Did I Ever Tell You About the Whale?, Or Measuring Technology Maturity. Charlotte, North Carolina: Information Age Publishing (2008)

22. Olechowski, A.L., Eppinger, S.D., Joglekar, N.: Technology Readiness Levels at 40: A Study of State-of-the-Art Use, Challenges, and Opportunities. MIT Sloan Research Paper No. 5127-15. Tech. rep., MIT (2015), http://dx.doi.org/10.2139/ssrn.2588524
23. Organisation for Economic Co-operation and Development (OECD): Inventory of Privacy-Enhancing Ttechnologies (PETs). Report DSTI/ICCP/REG(2001)1/FINAL, Working Party on Information Security and Privacy. Tech. rep. (2002), https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=dsti/iccp/reg%282001%291/final
24. Sauser, B.J., Verma, D., Ramirez-Marquez, J.E., Gove, R.: From TRL to SRL: The Concept of Systems Readiness Levels. In: Proc. Conference on Systems Engineering Research, Los Angeles, CA (2006)
25. Smith, J.D.: An Alternative to Technology Readiness Levels for Non-Developmental Item (NDI) Software. Tech. Rep. CMU/SEI-2004-TR-013, Software Engineerign Institute, Carnegie Mellon (2004)
26. Smith, J.D.: An Alternative to Technology Readiness Levels for Non-Developmental Item (NDI) Software. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences – HICSS '05 (2005)
27. Sweeney, L.: k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10(5), 557–570 (2002)